GAME DYNAMICS WITH LEARNING AND EVOLUTION OF UNIVERSAL GRAMMAR

W. GARRETT MITCHENER

Abstract.
We investigate a model of language evolution, based on population game dynamics with learning. Specifically, we examine the case of two genetic variants of universal grammar (UG), the heart of the human language faculty, assuming each admits two possible grammars. The dynamics are driven by a communication game. We prove using dynamical systems techniques that if the payoff matrix obeys certain constraints, then the two UGs are stable against invasion by each other, that is, they are evolutionarily stable. These constraints are independent of the learning process. Intuitively, if a mutation in UG results in grammars that are incompatible with the established languages, then it will die out because individuals with the mutation will be unable to communicate and therefore unable to realize any potential benefit of the mutation. An example for which the proofs do not apply shows that compatible mutations may or may not be able to invade, depending on the population's history and the learning process. These results suggest that the genetic history of language is constrained by the need for compatibility and that mutations in the language faculty may have died out or taken over depending more on historical accident than on any simple notion of relative fitness.

Key words. population game dynamics, replicator equation, language dynamical equation, learning, evolution, evolutionary stability

AMS subject classifications. 37N25, 92D15, 91F20

1. Introduction. Language is perhaps the most striking and specialized aspect of the human species. The ability of children to learn a language instinctively by listening to adult speech is especially remarkable, and many questions may be posed about the biological processes behind this ability [28, 3, 15, 20]. Using the framework of population game dynamics [6], this paper addresses the questions of whether genetic variation might exist in the language faculty, and under what circumstances can a mutation in the language faculty spread through a population. Mathematical models are extremely important in exploring the evolutionary history of human language, given the limited physical evidence [5, 14, 2, 8, 9, 29, 7, 22, 21]. The fundamental questions of how changes in the genetic code for language arise and propagate, in addition to being fascinating from the biological and philosophical perspectives, leads to interesting mathematics: The interaction of game dynamics and learning leads to a system of differential equations with rich behavior. A genetic variant is *evolutinarily stable* if the set of population states in which everyone has that variant are attracting. In this paper, we investigate a particular case of this dynamical system, and derive sufficient conditions for evolutionary stability of genetic variants of the language faculty.

Following prior work [24, 27, 31, 25, 30, 26, 13, 23], the benefit of language are represented by payoffs in a communication game, leading to reproductive success. However, genetic diversity in combination with learning adds extra complexity: The learning process is determined genetically, but the payoff comes from a strategy selected by the learning process. Thus, learning is an extra layer of indirection between the genome and the payoff. In the specific case of language, human beings are endowed with a set of innate hints and limitations known as *universal grammar* or *UG* [3, 28] that determine what languages are possible, and how children acquire a native language. UG may be thought of as a means for selecting a language, which in turn is a strategy for a communication game. That observation suggests the term *metastrategy* for traits such as UG which specify a strategy for selecting a strategy in a game.

The mathematical task at hand is to formulate a model of competition among UGs by adapting game dynamics to include metastrategies.

For such a mathematical model, a fundamental question is: Given a homogeneous population, if a small sub-population with a different UG is introduced, does the invasion die out, take over, or coexist? Initially, one might expect to find some measure of fitness such that the UG of greater fitness wins. However, as this paper shows, there is no such simple measure of fitness for metastrategies, because the payoff leading to fitness is not derived directly from a player's metastrategy.

This paper focuses on a restricted but non-trivial case where there are two UGs, each of which admits two grammars. There is considerable debate about exactly what parts of human intelligence are part of UG. For the purposes of this paper, a UG consists of a set of admissible languages and a learning algorithm that takes sample sentences and leads a child to acquire an admissible grammar. The main result is a mathematical statement of the following intuition. One might expect that if two UGs specify incompatible languages, then neither should be able to invade the other. Invaders confined to incompatible languages would die out not because their genetic endowment is less fit in any simple sense, but because they cannot realize the potential benefit of language if the majority of the population cannot communicate with them. For a range of parameter settings of the mathematical model, it is possible to have two UGs, each of which is stable against invasion by the other.

Furthermore, it is possible to have *accidental stability,* that is, two UGs, each of which can invade the other or not, depending on the dominant language. In this case, the population chooses a dominant language more or less at random, based on its initial state, and that accident determines whether or not a future invasion succeeds.

Section 2 formulates the language dynamical equation with multiple universal grammars. The resulting system of differential equations takes a payoff matrix and a learning matrix as parameters. From the general case, we restrict our attention to the case of two UGs with two grammars each. Section 3 derives conditions on the payoff matrix that imply that genetically homogeneous populations are stable against invasion by the other UG. These conditions are independent of the learning matrix and apply to any learning algorithm within the scope of the model. The conditions rigorously describe when a mutated UG is sufficiently incompatible with the existing UG to invade. Section 3.1 is a simplified proof for the case of two highly symmetric UGs, and Section 3.2 is the complete proof for general parameter settings. A brief discussion of these results, including an example of accidental stability, appears in Section 3.3. Finally, Section 4 draws some conclusions and indicates directions for further research.

2. The model. Following [12, 11, 16, 18, 19], we assume a large, well-mixed population, where all members have one of N genetically-determined UGs $U_1, U_2, \ldots, U_N$ and learn to speak one of n grammars $G_1, G_2, \ldots, G_n$. For simplicity, we assume that everyone speaks to everyone else, and that the lexicon is common to all speakers. In addition, reproduction will be modeled asexually, and we will ignore bilingualism.

The population as a whole is represented by variables $x_{j,K}$ representing the fraction of the population with U_K that speaks G_j. The sum of all segments of the population must total 100%, so $\sum_K \sum_j x_{j,K} = 1$. Since each $x_{j,K} \geq 0$, the set of all population states is a simplex. The fractions of the population with U_K, denoted y_K,

are defined as

$$y_K = \sum_{j=1}^{n} x_{j,K}. \tag{2.1}$$

For ease of notation, we also define variables for the fraction speaking G_j,

$$w_j = \sum_{K=1}^{N} x_{j,K}. \tag{2.2}$$

The dynamics are driven by a communication game. The payoff matrix is denoted B where $B_{i,j}$ is the payoff to a speaker of G_i when interacting with a speaker of G_j. One would expect that the maximum payoff occurs when both participants use the same grammar, so B ought to be diagonally dominant. As a concrete example, one could consider a communication game where $B_{i,j}$ is the probability that a speaker of G_i understands a random sentence spoken by a speaker of G_j. However, no particular form of B is required in this paper.

Learning is modeled by a stochastic matrix Q where $Q_{i,j,K}$ is the probability that a child of a speaker of G_i ends up speaking G_j, given that both have U_K. All children are assumed to learn some language, so for all i and K, $\sum_j Q_{i,j,K} = 1$. Since natural languages are generally passed on successfully, it is typical to assume that the diagonal entries $Q_{j,j,K}$ are close to 1, although that assumption is not required in this paper. In other papers on language dynamics [16, 11], the Q matrix is restricted to be constant. Here, that restriction is not necessary, and Q may in fact depend on x and t. We assume that parents always pass their UG to their children unchanged. Genetic mutation is assumed to be rare, and will be introduced through discrete invasion events, that is, perturbations of the game dynamics.

As in [18], the language dynamical equation for multiple UGs is

$$\begin{aligned} F &= Bw, \\ \phi &= w^T F = \sum_j w_j F_j, \\ \dot{x}_{j,K} &= \sum_{i=1}^{n} F_i x_{i,K} Q_{i,j,K} - \phi x_{j,K} \\ &= x_{j,K}(F_j Q_{j,j,K} - \phi) + \sum_{i \neq j} F_i x_{i,K} Q_{i,j,K}. \end{aligned} \tag{2.3}$$

Each entry F_j of the vector F represents the average payoff to a speaker of G_j, given that each G_j is spoken by a fraction w_j of the population. The base reproductive rate for speakers of G_j is tied directly to its payoff F_j. The average fitness over the whole population is given by ϕ. The second form of the equation for $\dot{x}_{j,K}$ is perhaps the easiest to explain, as it shows the resemblance to the standard replicator model [6]. Roughly, the change in $x_{j,K}$ is determined by how much the fitness (or basic reproductive rate) of G_j exceeds the population average ϕ. The remaining terms involving Q represent imperfect learning.

The dynamical system (2.3) does not account for spatial and social structure of the population, and the representation of learning as a stochastic matrix ignores much of the complexity of true language acquisition. However, it is simple enough that one can hope for significant theoretical results. Alternative models include greater detail

at the expense of additional complexity, and the analysis of such simulations is often limited to statistical results [1, 2, 10].

For the remainder of the paper, we assume that there are two UGs with two grammars each. Thus, U_1 admits G_1 and G_2 and U_2 admits G_3 and G_4. We require $x_{j,K}$ to be fixed at 0 if U_K does not admit G_j, and the necessary slots in Q must also be zero, so for example $Q_{1,3,1}$ is zero because children with U_1 cannot learn G_3.

Determining all possible behaviors of even the two UG system in general is an extremely difficult problem, so we focus only on those population states where initially everyone has the same UG and the other UG attempts to invade. Section 3 poses the question mathematically by examining population states where $y_K = 1 - \varepsilon$. An *attracting set* [4] is a closed invariant subset of the phase space surrounded by a neighborhood in which every trajectory tends to the set in forward time. If the population tends back to $y_K = 1$, then the set of all states with $y_K = 1$ forms an attracting set, and U_K is stable against invasion by the other UG. Thus the dynamics of the y_K variables will be crucial. Since Q is row stochastic, the expression for $\dot{y}_K$ simplifies considerably, leaving just

$$\dot{y}_K = \sum_i F_i x_{i,K} - \phi y_K. \tag{2.4}$$

Remarkably, the Q matrix disappears. The learning process still influences the dynamics of y_K in that it steers the $x_{i,K}$. However, as will be shown in Section 3, there is the possibility that the overall behavior of the y_K's, and hence the evolutionary stabilities of U_1 and U_2, can sometimes be determined without reference to Q. In such cases, the B matrix alone determines whether a UG is stable, and any learning algorithm, even one that depends on x and t, will yield the same result.

The curious reader is invited to read Chapter 5 of [17] for additional results in more specific cases.

2.1. New coordinates. To analyze (2.3), we begin by changing coordinates so that the behavior of the y_K's is more readily apparent. The original variables $x_{j,K}$ will be called *simplex coordinates,* and the phase space will be drawn as a pyramid as in Figure 2.1. The corner points $X_{j,K}$ represent extreme populations where $x_{j,K} = 1$ and the other x's are 0. The following new variables will be called *box coordinates*:

$$\begin{aligned} r &= \frac{x_{2,1} - x_{1,1}}{x_{2,1} + x_{1,1}}, -1 \le r \le 1, \\ s &= \frac{x_{4,2} - x_{3,2}}{x_{4,2} + x_{3,2}}, -1 \le s \le 1, \\ z &= x_{1,1} + x_{2,1} = y_1, 0 \le z \le 1. \end{aligned} \tag{2.5}$$

The balance between G_1 and G_2 is represented by r. Likewise, the balance between G_3 and G_4 is represented by s. Since $z = y_1$, it represents the fraction of the population with U_1. These three pieces of information are enough to identify all possible

population states. The reverse change of coordinates is:

$$\begin{aligned} x_{1,1} &= \left(\frac{1-r}{2}\right) z, \\ x_{2,1} &= \left(\frac{1+r}{2}\right) z, \\ x_{3,2} &= \left(\frac{1-s}{2}\right) (1-z), \\ x_{4,2} &= \left(\frac{1+s}{2}\right) (1-z). \end{aligned} \tag{2.6}$$

It is worth noting that the change of coordinates is singular: It expands the simplex in $x_{j,K}$ coordinates into a box in (r, s, z) by blowing up the edges $X_{1,1}X_{2,1}$ and $X_{3,2}X_{4,2}$ into squares, as illustrated in Figure 2.1.

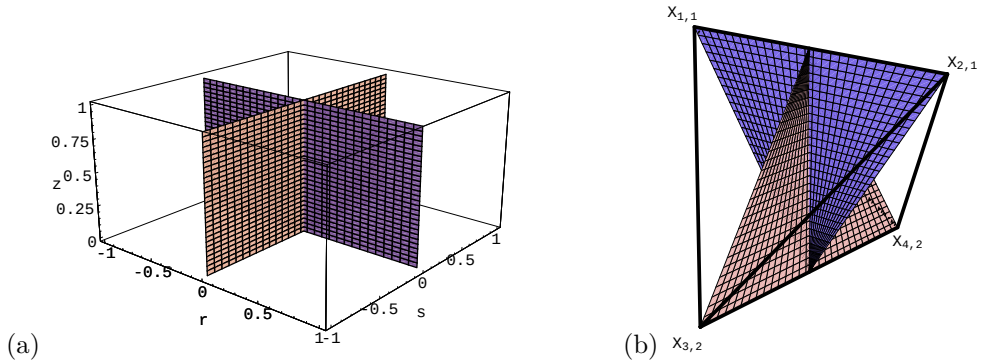


FIG. 2.1. *The singular change of coordinates. (a): The planes given by $r = 0$ and $s = 0$ in (r, s, z) or box coordinates. (b): The corresponding sets in simplex coordinates. Note that the top edge of the simplex, where $y_1 = 1$, corresponds to a square in box coordinates. Similarly for the bottom edge, where $y_1 = 0$.*

3. Stability conditions. The learning algorithm does not explicitly appear in the time derivative of the size of the sub-population with a given U_K, shown in (2.4). This observation motivates a series of calculations that yields sufficient conditions for evolutionary stability. These six inequalities may be easily interpreted in a number of special cases, thus building mathematical intuition for when evolutionary stability occurs.

A simple calculation using (2.4) shows that $\dot{y}_K = 1$ if $y_K = 0$, so for each K the set of points where $y_K = 1$ is closed and invariant. These are the upper and lower edges of the simplex in Figure 2.1. To show that these edges are attracting, we will divide the simplex into three regions, one in which all populations tend toward the top edge where U_1 takes over, one in which all populations tend toward the bottom edge where U_2 takes over, and an intermediate region. To illustrate the argument, we first analyze a case with a highly symmetric payoff matrix B, then generalize the argument to the case of general B.

3.1. Illustration of the null-cline argument in the case of permutation symmetry. To illustrate the argument, consider the case where all the grammars are

interchangeable as far as the payoff they generate, so

$$B = \begin{pmatrix} 1 & a & a & a \\ a & 1 & a & a \\ a & a & 1 & a \\ a & a & a & 1 \end{pmatrix}. \tag{3.1}$$

This symmetry simplifies the dynamical system considerably. There are two universal grammars, so we have two variables of interest, $y_1 = x_{1,1} + x_{2,1}$ and $y_2 = x_{3,2} + x_{4,2}$. Since $y_1 + y_2 = 1$, we need only analyze the behavior of y_1. The following proposition, describes how the limiting behavior of y_1 is largely determined by the initial population state.

PROPOSITION 3.1. *The simplex contains two trapping regions which are independent of the Q matrix: Trajectories for which $y_1(0) > 2/3$ tend to $y_1 = 1$, and trajectories for which $y_1(0) < 1/3$ tend to $y_1 = 0$. In the region in between, the Q matrix influences whether y_1 approaches 1 or 0.*

Proof. We work in box coordinates. Observe that $y_1 = z$, from which we may calculate that

$$\dot{z} = \dot{y}_1 = \frac{1}{2}(1-a)(1-z)z(-1-s^2(1-z)+(2+r^2)z). \tag{3.2}$$

To find the Q-independent trapping regions, we first look for the z null-clines. These are the sets of points for which $\dot{z} = 0$. From (3.2) it is clear that $\dot{z} = 0$ if and only if $z = 0$, $z = 1$, or $z = h(r, s)$ where

$$h(r,s) = \frac{1+s^2}{2+s^2+r^2}.$$

The first two cases are the upper and lower edges of the simplex, and the third is a surface in the middle. See Figure 3.1. Thus, $\dot{z}$ is of one sign above the surface and the opposite sign below. Looking at the vertical line given by $r = 0$ and $s = 0$, we have $\dot{z} = -\frac{1}{2}z(-1+z)(-1+2z)$ which is positive for $z > 1/2$ and negative for $z < 1/2$.

Therefore, the overall picture is that the simplex decomposes into upper and lower trapping regions and a boundary region in the middle. If a trajectory starts above the topmost point of the z null-cline, then $\dot{y}_1 > 0$, which means y_1 will increase over time until it reaches $y_1 = 1$. Likewise, any trajectory that starts below the bottommost point of the surface will continue downward until $y_1 = 0$.

To find the topmost and bottommost points, observe that the surface is saddle shaped, so the extrema will appear on the boundaries. Looking on the faces of the box given by $s = \pm 1$, we have $h(r, \pm 1) = 2/(3 + r^2)$ which has a maximum at $h(0, \pm 1) = 2/3$ and minima on the edges at $h(\pm 1, \pm 1) = 1/2$. Likewise, looking on the surfaces given by $r = \pm 1$, we have $h(\pm 1, s) = 1 - 2/(3 + s^2)$ which has a minimum at $h(\pm 1, 0) = 1/3$. Furthermore, it has maxima on the edges at $h(\pm 1, \pm 1) = 1/2$. Therefore, if either universal grammar holds a 2/3 majority of the population, it will eventually take over regardless of the Q matrix. □

In the boundary region near the z null-cline, many orbits obey the simple rule that if they start above the surface, they approach $y_1 = 1$ and if they start below, they approach $y_1 = 0$. However, orbits may pass through the surface horizontally, thereby starting above it but converging to $y_1 = 0$ or starting below it but converging to $y_1 = 1$. For example, Figure 3.2 shows the values of y_1 and y_2 starting from a point just above the z null-cline for which $y_1 \to 1$. However, a nearby initial condition

produces the trajectories in Figure 3.3, where y_1 passes horizontally through the z null-cline and turns downward.

The actual surface dividing orbits that go to $y_1 = 1$ from those that go to $y_1 = 0$ depends on Q. For example, it might be the stable manifold of a saddle point in the middle of the simplex, or perhaps something more complicated.

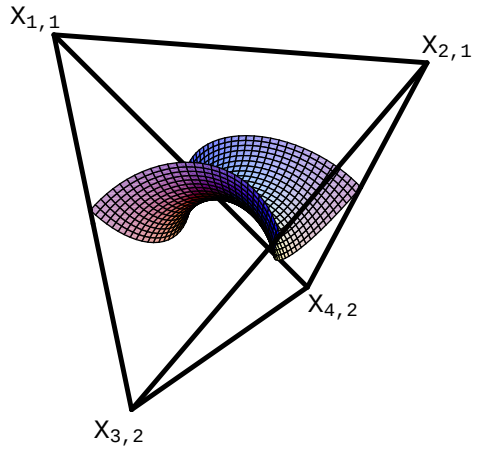
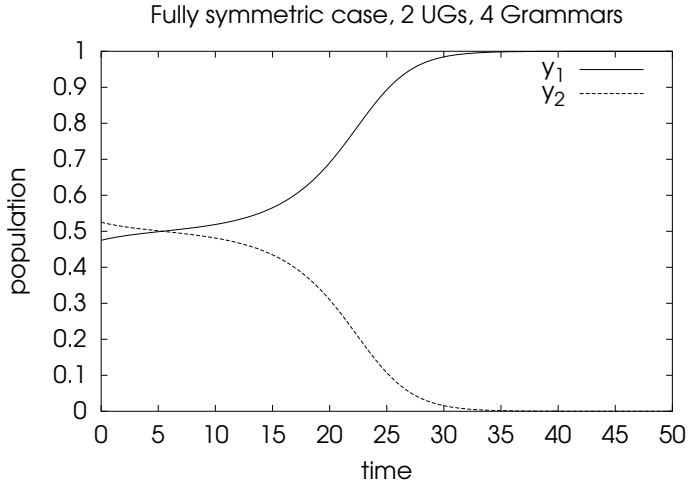


FIG. 3.1. *The upended simplex and z null-cline.*



$$a = 1/10, Q_{i,j,1} = \begin{pmatrix} 0.7 & 0.3 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, Q_{i,j,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$

FIG. 3.2. *Trajectories starting from* $x_{1,1} = 57/160 = 0.35625, x_{2,1} = 19/160 = 0.11875, x_{3,2} = x_{4,2} = 21/80 = 0.2625$ *which lies just above the z null-cline.*

The proposition implies that in this case, the learning algorithms employed by the two universal grammars and specified by Q are largely irrelevant to determining which universal grammar takes over the population. As long as a certain majority of

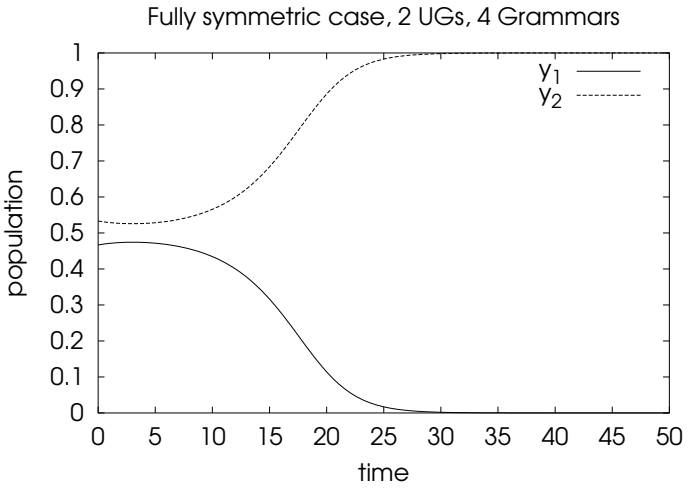


FIG. 3.3. *Trajectories starting from* $x_{1,1} = 7/20 = 0.35, x_{2,1} = 7/60 = 0.11\bar{6}, x_{3,2} = x_{4,2} = 4/15 = 0.2\bar{6}$. *Here, the trajectory passes through the* z *null-cline, and* y_1 *increases, reaches a maximum, then turns downward and tends to* 0. *The parameters are the same as in Figure 3.2.*

the population uses one universal grammar, it will be evolutionarily stable.

3.2. Null-cline argument in the general case. In this section, we extend the null-cline argument of Section 3.1 to the case of fully general B and Q matrices. We first find sufficient conditions under which the null-cline does not intersect the top and bottom edges of the simplex. This guarantees that there are regions that lie completely above or below it. Second, we find a condition that implies that trajectories above the surface move upward, and those below it move downward, thereby assuring that the top and bottom edges are attracting sets.

We will allow the B matrix to be completely general, with entries $b_{i,j}$. In what follows, more concise expressions result if the following parameters are used instead of the entries of B:

$$\alpha_0 = \frac{1}{2}(b_{11} + b_{12} + b_{21} + b_{22}) \qquad \alpha_1 = \frac{1}{2}(b_{11} - b_{12} - b_{21} + b_{22})$$
$$\alpha_2 = \frac{1}{2}(b_{11} + b_{12} - b_{21} - b_{22}) \qquad \alpha_3 = \frac{1}{2}(b_{11} - b_{12} + b_{21} - b_{22})$$
$$\beta_0 = \frac{1}{2}(b_{13} + b_{14} + b_{23} + b_{24}) \qquad \beta_1 = \frac{1}{2}(b_{13} - b_{14} - b_{23} + b_{24})$$
$$\beta_2 = \frac{1}{2}(b_{13} + b_{14} - b_{23} - b_{24}) \qquad \beta_3 = \frac{1}{2}(b_{13} - b_{14} + b_{23} - b_{24})$$
$$\gamma_0 = \frac{1}{2}(b_{31} + b_{32} + b_{41} + b_{42}) \qquad \gamma_1 = \frac{1}{2}(b_{31} - b_{32} - b_{41} + b_{42})$$
$$\gamma_2 = \frac{1}{2}(b_{31} + b_{32} - b_{41} - b_{42}) \qquad \gamma_3 = \frac{1}{2}(b_{31} - b_{32} + b_{41} - b_{42})$$
$$\delta_0 = \frac{1}{2}(b_{33} + b_{34} + b_{43} + b_{44}) \qquad \delta_1 = \frac{1}{2}(b_{33} - b_{34} - b_{43} + b_{44})$$
$$\delta_2 = \frac{1}{2}(b_{33} + b_{34} - b_{43} - b_{44}) \qquad \delta_3 = \frac{1}{2}(b_{33} - b_{34} + b_{43} - b_{44})$$

We will work in box coordinates again, as defined in (2.5). After simplification,

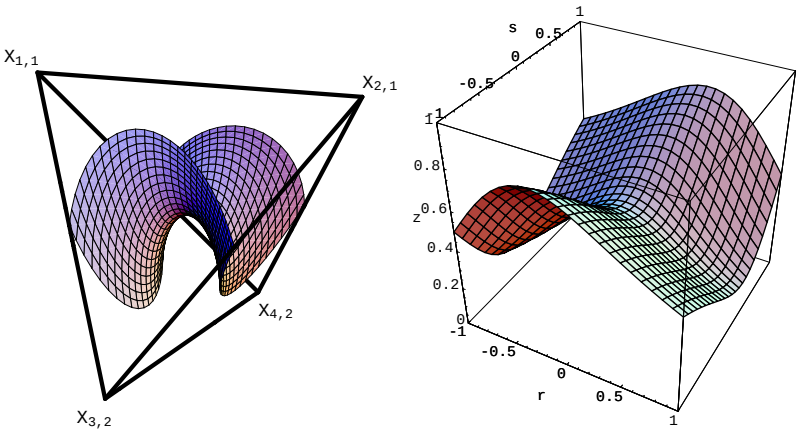
$$\dot{z} = \frac{1}{4}(-1+z)zg(r,s,z), \tag{3.3}$$

where

$$\begin{aligned} g(r,s,z) =& 2\Big(-\beta_0 + \delta_0 + r\beta_2 + s(\beta_3 - \delta_2 - \delta_3) - z(\alpha_0 - \beta_0 - \gamma_0 + \delta_0) \\ & - rs\beta_1 + rz(\alpha_2 + \alpha_3 - \beta_2 - \gamma_3) - sz(\beta_3 + \gamma_2 - \delta_2 - \delta_3) \\ & + s^2\delta_1 + rsz(\beta_1 + \gamma_1) - r^2z\alpha_1 - s^2z\delta_1 \Big) \end{aligned} \tag{3.4}$$

The factorized form (3.3) shows that there are three z null-clines: the top ($z = 1$), the bottom ($z = 0$), and the surface determined by $g(r, s, z) = 0$. This surface will be called the interior z null-cline. The goal of this section is to determine sufficient conditions on B such that the interior z null-cline creates trapping regions around the top and bottom edges of the simplex.

3.2.1. Step 1: The non-intersection constraints. The first condition is that the interior z null-cline must not touch the top and bottom, implying that there is some space between the vertical extrema of the surface and the top and bottom edges of the simplex. See Figure 3.4 for an example.



$$B = \begin{pmatrix} 1 & \frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & \frac{1}{4} & 1 \end{pmatrix}$$

FIG. 3.4. *The interior z null-cline in simplex coordinates (left) and box coordinates (right). The asymmetric B matrix used to generate these pictures is as shown.*

PROPOSITION 3.2. *Suppose the following expressions are all strictly positive:*

$$\begin{aligned}
\nu_1 &= 4\delta_1(\delta_0 - \beta_0 + \beta_2) - (\beta_1 - \beta_3 + \delta_2 + \delta_3)^2 \\
\nu_2 &= 4\delta_1(\delta_0 - \beta_0 - \beta_2) - (\beta_1 + \beta_3 - \delta_2 - \delta_3)^2 \\
\nu_3 &= 4\alpha_1(\alpha_0 - \gamma_0 + \gamma_2) - (\alpha_2 + \alpha_3 + \gamma_1 - \gamma_3)^2 \\
\nu_4 &= 4\alpha_1(\alpha_0 - \gamma_0 - \gamma_2) - (\alpha_2 + \alpha_3 - \gamma_1 - \gamma_3)^2
\end{aligned} \tag{3.5}$$

and also that

$$\beta_0 - \delta_0 > 0 \text{ and } \gamma_0 - \alpha_0 > 0. \tag{3.6}$$

Then, the interior z null-cline lies strictly between the top and bottom of the simplex at a strictly positive distance from each.

Proof. The mathematical formulation of the conclusion in box coordinates is that if $-1 \le r \le 1$ and $-1 \le s \le 1$, then $g(r, s, 0) \neq 0$ and $g(r, s, 1) \neq 0$. We proceed by proving that g is of one sign on the bottom plane $z = 0$, and also of one sign on the top plane $z = 1$. We must add the technical assumptions that α_1 and δ_1 are nonzero to eliminate some degenerate cases that would cause division by zero in what follows. These assumptions are harmless, as ν_1 and ν_2 cannot possibly be positive if $\delta_1 = 0$, and ν_3 and ν_4 cannot possibly be positive if $\alpha_1 = 0$.

For the bottom, we are interested in $g(r, s, 0)$, which happens to be a quadratic form in r and s, so the equation $g(r, s, 0) = 0$ must define a conic section in the plane $z = 0$. To classify it, we complete the square in r and s and change variables to ρ and σ so as to put it in a standard form:

$$\begin{aligned}
r &= \frac{2\beta_2\delta_1 + \beta_1(\beta_3 - \delta_2 - \delta_3 - 2\delta_1\rho)}{\beta_1^2}. \\
s &= \frac{\beta_2}{\beta_1} - \rho + \sigma.
\end{aligned}$$

With these new variables, the equation $g(r, s, 0) = 0$ becomes

$$-\frac{2(\beta_0\beta_1^2 - \beta_1^2\delta_0 - \beta_2^2\delta_1 + \beta_1\beta_2(-\beta_3 + \delta_2 + \delta_3))}{\beta_1^2} - 2\delta_1\rho^2 + 2\delta_1\sigma^2 = 0,$$

which is the form of a hyperbola in ρ and σ. If it happens that $\beta_1 = 0$, then the above change of variables is not applicable, and the curve turns out to be a parabola. Either way, if we want to specify that the interior z null-cline does not touch the bottom, it is sufficient to require that the curve specified by $g(r, s, 0) = 0$ lies outside the square given by $-1 \le r \le 1$ and $-1 \le s \le 1$. For a picture of an example of this curve, see Figure 3.5. That constraint is equivalent to requiring the expression $g(r, s, 0)$ to be of one sign on the sides of the square. To avoid having two separate cases ($g > 0$ or $g < 0$), we transform the constraint by dividing $g(r, s, 0)$ by the coefficient of s^2 and requiring the resulting expression $g_2(r, s)$ to be positive on the sides of the square:

$$g_2(r, s) = s^2 + \frac{-\beta_0 + r\beta_2 + \delta_0}{\delta_1} - \frac{s(r\beta_1 - \beta_3 + \delta_2 + \delta_3)}{\delta_1}. \tag{3.7}$$

(See Figure 3.6.)

Note that on the sides where $r = \pm 1$, the expressions $g(\pm 1, s)$ are quadratic in s, and open upward because the coefficient of s^2 is 1. So, to guarantee that $g_2(r, s) > 0$

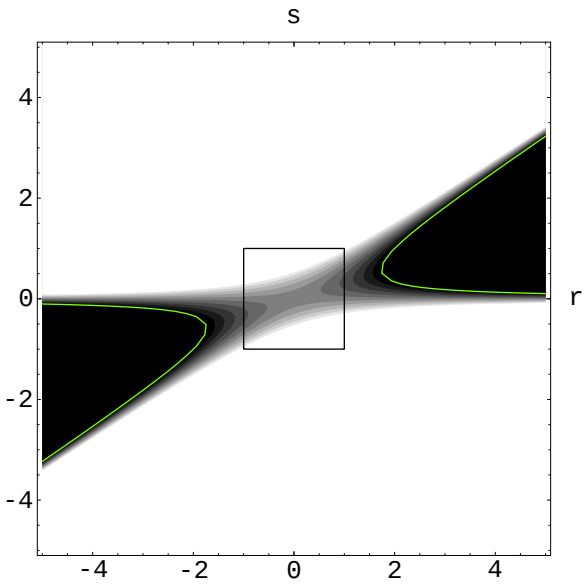


FIG. 3.5. *Contour plot of $g(r, s, 0)$. The light hyperbola is $g(r, s, 0) = 0$, that is, the curve where the interior z null-cline intersects the plane $z = 0$. For other points, dark shading indicates a negative value of $g(r, s, 0)$ and lighter shading indicates a positive value. The square is the bottom face of the phase space. See Figure 3.4 for the particular B used in this illustration. This is where we have to think outside the box.*

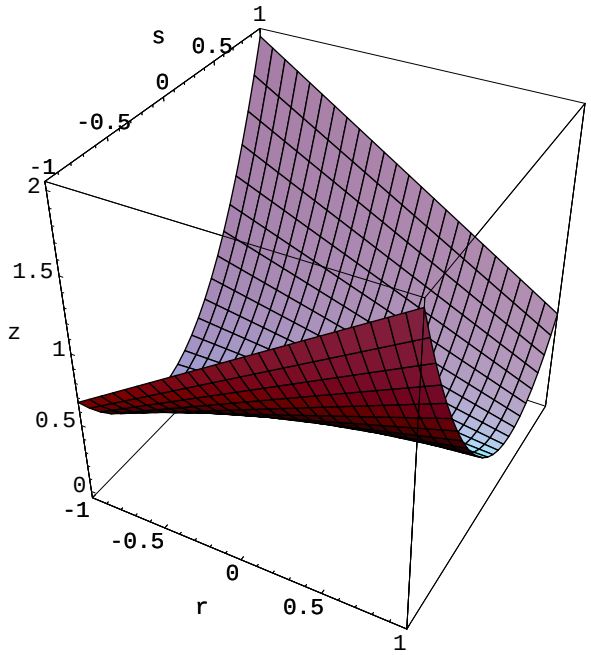


FIG. 3.6. *Plot of $g_2(r, s)$, which differs by a constant factor from $g(r, s, 0)$, using the same B as in Figure 3.4. The region shown is the phase space in box coordinates. Observe that the surface intersects with two faces of the phase space in parabolas, and with the other in lines.*

on these two sides, it suffices to require that the minima of $g_2(\pm 1, s)$ be positive. The minimum of a general quadratic function $x^2 + ax + b$ is $b - a^2/4$, so the exact constraints are

$$\min_s g_2(1,s) = -\frac{-4(-\beta_0+\beta_2+\delta_0)\delta_1 + (\beta_1-\beta_3+\delta_2+\delta_3)^2}{4\delta_1^2} > 0,$$
$$\min_s g_2(-1,s) = -\frac{4(\beta_0+\beta_2-\delta_0)\delta_1 + (\beta_1+\beta_3-\delta_2-\delta_3)^2}{4\delta_1^2} > 0.$$

Both denominators are square, so only the numerators matter in satisfying the inequalities. We therefore simplify the constraints to the first two inequalities in the statement of the proposition, namely $\nu_1 > 0$ and $\nu_2 > 0$.

Observe that if these constraints are satisfied, then $g_2(r,s) > 0$ on all four corners of the square. With that observation, the sides where $s = \pm 1$ are easy to check, as $g_2(r, \pm 1)$ is a linear function of r, and it is therefore enough to require that $g_2(r,s) > 0$ on the corners. In summary, if $\nu_1 > 0$ and $\nu_2 > 0$, then g is of one sign on all four sides of the bottom square in box coordinates, and therefore, the z null-cline does not intersect the bottom of the simplex.

The constraint that the interior z null-cline cannot intersect with the top of the simplex can be enforced by imposing a second set of inequalities similar to those discovered above. Again, the equation for where the null-cline intersects $z = 1$ is $g(r,s,1) = 0$ which defines a hyperbola in the plane $z = 1$ in terms of r and s. To specify that the null-cline does not touch the top edge of the simplex, it suffices to require that this hyperbola lie outside the square in box coordinates given by $-1 \le r \le 1$ and $-1 \le s \le 1$. As before, we ensure this by requiring $g(r,s,1)$ to be of one sign on all four sides of the square. Equivalently, we define $g_3(r,s)$ to be $g(r,s,1)$ divided by the coefficient of r^2, and require $g_3(r,s)$ to be positive on all four sides of the square. The expression for g_3 is

$$g_3(r,s) = r^2 + \frac{\alpha_0 - \gamma_0 + s\gamma_2}{\alpha_1} - \frac{r(\alpha_2+\alpha_3+s\gamma_1-\gamma_3)}{\alpha_1}. \tag{3.8}$$

Furthermore, $g_3(r, \pm 1)$ are monic quadratic functions of r, so it suffices to require that their minima be positive, which yields

$$\min_r g_3(r,1) = -\frac{-4\alpha_1(\alpha_0-\gamma_0+\gamma_2) + (\alpha_2+\alpha_3+\gamma_1-\gamma_3)^2}{4\alpha_1^2} > 0,$$
$$\min_r g_3(r,-1) = -\frac{4\alpha_1(-\alpha_0+\gamma_0+\gamma_2) + (\alpha_2+\alpha_3-\gamma_1-\gamma_3)^2}{4\alpha_1^2} > 0.$$

As before, the denominators are all square, so only the numerators matter, and the constraints reduce to $\nu_3 > 0$ and $\nu_4 > 0$. These imply that g is of one sign on the top of the phase space in box coordinates, and that the z null-cline does not intersect with the top of the simplex.

The final two constraints in the statement of the proposition are there to ensure that the interior null-cline lies inside the simplex rather than completely above or below it, and are derived as follows. Choose $\bar{z}$ such that $g(0,0,\bar{z}) = 0$, that is, the point at which the null-cline intersects the vertical line given by $r = 0$ and $s = 0$:

$$\bar{z} = \frac{\delta_0 - \beta_0}{\alpha_0 - \gamma_0 + \delta_0 - \beta_0}. \tag{3.9}$$

A short calculation proves that the constraints $\delta_0 - \beta_0 > 0$ and $\alpha_0 - \gamma_0 > 0$ imply $0 < \bar{z} < 1$, which guarantees that the null-cline lies completely inside the simplex.

It turns out that $g(r, s, z) = 0$ can actually be solved in terms of z, and the resulting solution $z = h(r, s)$ is the quotient of two polynomials in r and s. Under the constraints derived in this proposition, h must be bounded for $-1 \leq r \leq 1$ and $-1 \leq s \leq 1$, which means its denominator never vanishes. Therefore, h is continuous, and since the region of interest for r and s is a closed square, h actually takes on its extreme values. It follows that there is a strictly positive distance between the null-cline and the top and bottom of the simplex. □

3.2.2. Step 2: Direction of the vector field. Now we must show that $\dot{z} > 0$ above the interior z null-cline, and $\dot{z} < 0$ below it. This claim implies that trajectories that pass above the uppermost point on the null-cline continue to rise, and those that pass below the lowermost point continue to fall, thereby establishing the existence of the two trapping regions.

PROPOSITION 3.3. *Assume that the interior z null-cline is strictly between the top and bottom edges of the simplex, and that $\delta_0 - \beta_0 > 0$ and $\alpha_0 - \gamma_0 > 0$. Then there are trapping regions above and below the null-cline.*

Proof. The null-clines are by definition the set of points where $\dot{z} = 0$, so in regions between them, $\dot{z}$ is of one sign. It therefore suffices to show that for some point above the interior null-cline, $\dot{z} > 0$, and for some point below it, $\dot{z} < 0$. Consider the vertical line given by $r = 0$ and $s = 0$, as illustrated in Figure 3.7. Along this line,

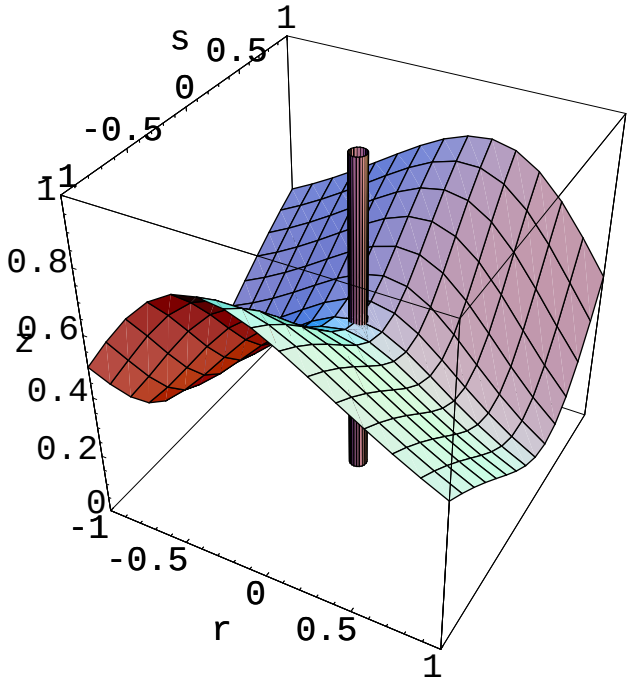
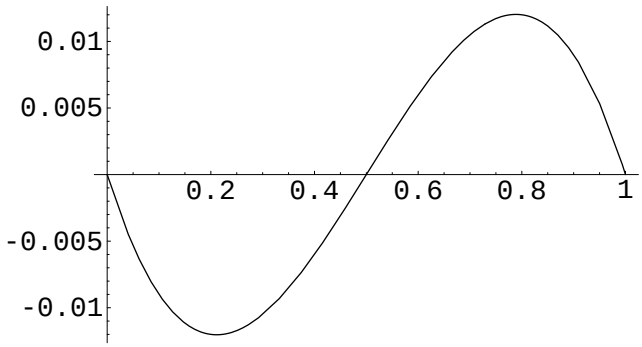


FIG. 3.7. *The z null-cline in box coordinates with the line $r = 0, s = 0$ indicated by a bar. The B matrix used for this picture is the same as in Figure 3.4*

$$\dot{z}|_{r=0,s=0} = \frac{1}{2}(-1 + z)z(-\beta_0 + z(-\alpha_0 + \beta_0 + \gamma_0 - \delta_0) + \delta_0).$$

That is, $\dot{z}$ is a cubic function $f(z)$ along this vertical line, as in Figure 3.8. We need



$$f(z) = \dot{z}|_{r=0,s=0}$$

FIG. 3.8. *The vertical component of the vector field along the central vertical line in Figure 3.7*

only require that $f'(0) < 0$ to ensure that $\dot{z}$ is negative below the null-cline and positive above it, which is equivalent to the inequality

$$f'(0) = \frac{\beta_0 - \delta_0}{2} < 0. \tag{3.10}$$

Equivalently, we may require that $f'(1) < 0$, which yields the inequality

$$f'(1) = \frac{\gamma_0 - \alpha_0}{2} < 0. \tag{3.11}$$

Both of these inequalities follow immediately from the hypotheses. □

3.3. Discussion. What makes these two propositions possible is the fact that Q is row-stochastic, so that it disappears in $\dot{y}_K$ in (2.4). The same would happen if Q depended on t or x. The trapping regions described in Propositions 3.2 and 3.3 are also independent of Q. Thus, for any collection of grammars that satisfies the hypotheses of these propositions, as long as one of the universal grammars has a sufficiently large majority of the population, it will take over no matter what the learning algorithm, static or dynamic. Although these propositions do not rule out the possibility of stable coexistence, they do provide conditions under which homogeneous populations are stable against invasion by the other UG.

Here is some intuition concerning the constraints in Proposition 3.2. Let us consider ν_1. If the greatest payoff occurs when two people with the same language interact, then the payoff matrix B should be diagonally dominant, which implies that $\delta_1 > 0$. The parameter β_0 is the sum of payoffs to individuals with U_1 when speaking to individuals U_2. The parameter δ_0 is the sum of payoffs to individuals with U_2 when speaking to others with U_2. So roughly, $\nu_1 > 0$ means that $\delta_0 - \beta_0$ is large, which means that U_2 receives a greater payoff when interacting with U_2 than U_1 does. That agrees with the concept of evolutionary stability.

We may also examine some more specific forms of the B matrix to see what payoffs lead to stable UGs. For example, consider the question of two very different UGs, where G_1 and G_2 do not communicate well with G_3 and G_4. The payoff matrix

for such a situation might look like this:

$$B_{\text{diff}} = \begin{pmatrix} c & a & \varepsilon & \varepsilon \\ a & c & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & c & a \\ \varepsilon & \varepsilon & a & c \end{pmatrix}. \tag{3.12}$$

where c is relatively large, ε is small, and a is in between. For a picture, see Figure 3.9. The constraints simplify greatly in this case:

$$\nu_1 = \nu_2 = \nu_3 = \nu_4 = 4(c-a)(c+a-2\varepsilon),$$

$$\alpha_0 - \gamma_0 = \delta_0 - \beta_0 = a + c - 2\varepsilon.$$

Clearly, if ε is small enough, then all six constraints are positive. Therefore, the two UGs are stable against invasion by each other no matter what their learning processes are.

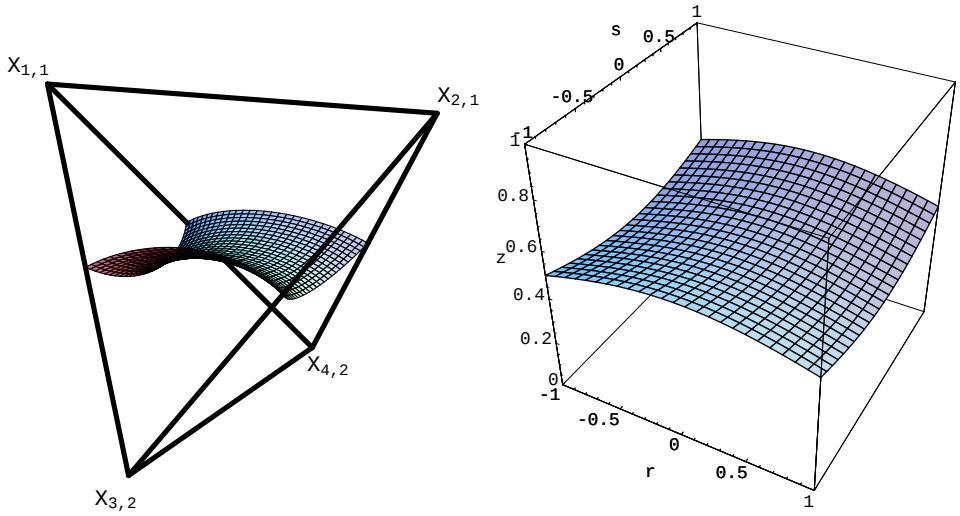


FIG. 3.9. *Null-cline for the case of two very different UGs, using the payoff matrix* B_{diff} *with* $c = 1$, $a = 1/2$, *and* $\varepsilon = 1/8$. *Left: Phase space in simplex coordinates. Right: Phase space in box coordinates.*

On the other hand, consider the case of two very similar UGs, where $G_1 \approx G_3$ and $G_2 \approx G_4$. The payoff matrix for this example might look like this:

$$B_{\text{sim}} = \begin{pmatrix} c & a & (1-\varepsilon)c & (1-\varepsilon)a \\ a & c & (1-\varepsilon)a & (1-\varepsilon)c \\ (1-\varepsilon)c & (1-\varepsilon)a & c & a \\ (1-\varepsilon)a & (1-\varepsilon)c & a & c \end{pmatrix}. \tag{3.13}$$

The constraints simplify to

$$\nu_1 = \nu_2 = \nu_3 = \nu_4 = -(c-a)^2 - 2(a^2 + 2ac - 3c^2)\varepsilon - (a-c)^2\varepsilon^2,$$

and

$$\alpha_0 - \gamma_0 = \delta_0 - \beta_0 = (a+c)\varepsilon.$$

If ε is small enough, then ν_1, ν_2, ν_3 and ν_4 are dominated by $-(c-a)^2$ which is negative, implying that the null-cline might intersect with the top and bottom of the phase space as illustrated in Figure 3.10. Propositions 3.2 and 3.3 do not apply, so it is possible that one UG might be able to invade the other, and the learning process is critical to understanding the long-term behavior of the system.

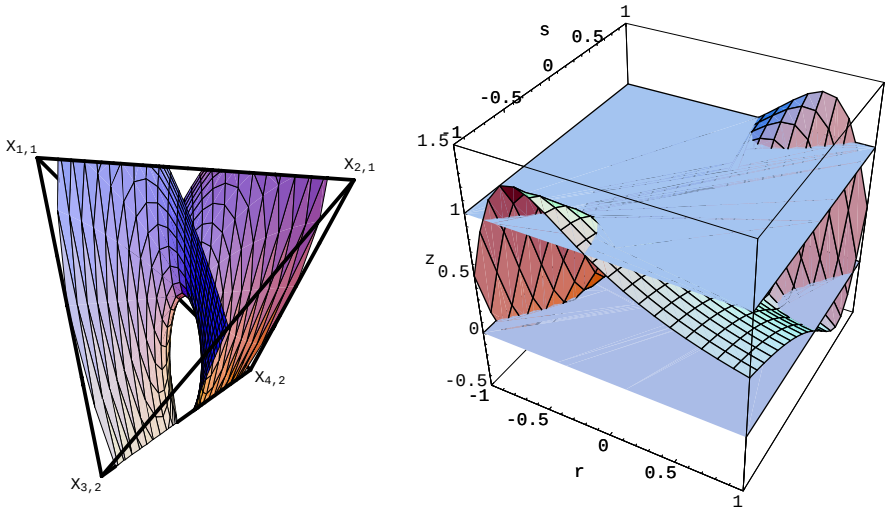
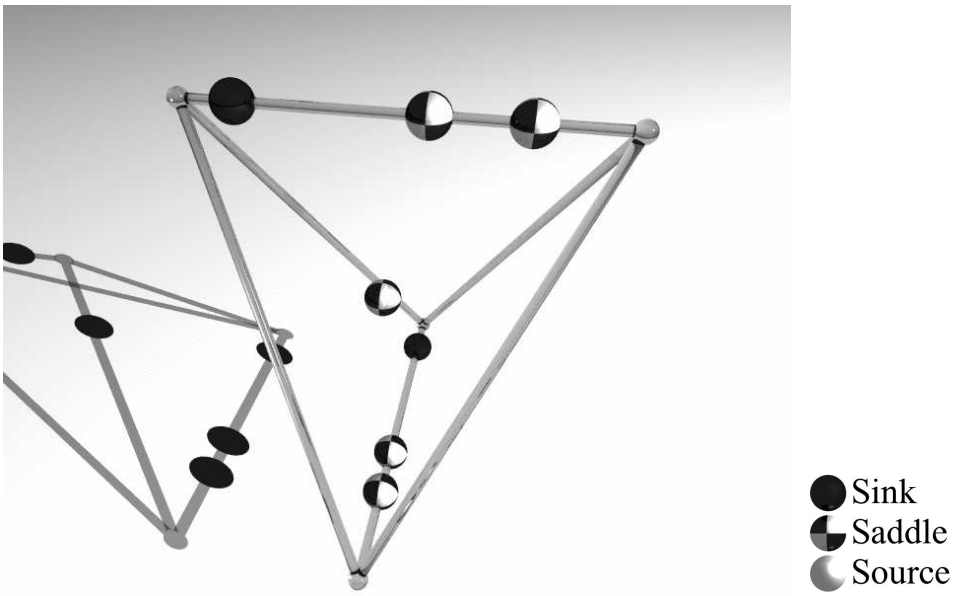


FIG. 3.10. *Null-cline for the case of two very similar UGs, using the payoff matrix* B_{sim} *with* $c = 1$, $a = 1/2$, *and* $\varepsilon = 1/32$. *Left: Phase space in simplex coordinates. Right: Phase space in box coordinates, with planes* $z = 0$ *and* $z = 1$ *indicated. Note that the null-cline intersects with these planes within the phase space.*

The case of similar grammars leads to a remarkable situation called *accidental stability,* illustrated by the 3-D phase portrait in Figure 3.11. This is a case of two similar UGs where $G_1 \approx G_3$ and $G_2 \approx G_4$, so it is no surprise that the two UGs can invade one another. What is surprising is the mechanism. Consider an initial population whose members all have U_1. These states are all on the top edge of the simplex and remain on that line unless subject to an external perturbation. Such a population will tend to one of the two fixed points on the top edge, one of which is dominated by G_1 and the other by G_2. The fixed point dominated by G_1 is a stable sink, and if U_2 tries to invade that population, it will fail. However, the fixed point dominated by G_2 is a saddle, and if U_2 tries to invade that population, the invasion succeeds, U_2 takes over completely, and the population tends to the sink on the bottom edge of the simplex dominated by G_4. The initial state of the all-U_1 population determines whether a later invasion by U_2 succeeds or not, and that initial state is essentially random. Hence, this instance of the language equation is sensitive to historical accidents.

4. Conclusion. For a range of values of the payoff matrix in the communication game, genetically homogeneous populations are evolutionarily stable. Roughly, a large difference in the languages admitted by two UGs causes them to be stable against invasion by each other. This fact has significant consequences for the evolution of UG: The benefits of communicating with the rest of the population limit evolution to innovations that are fairly compatible with the existing UG. Other potentially beneficial mutations are likely to die out before their benefits can be realized. In the



$$B = \begin{pmatrix} 1 & 0.2 & 0.99 & 0.1 \\ 0.2 & 1 & 0.1 & 0.99 \\ 0.99 & 0.1 & 1 & 0.2 \\ 0.1 & 0.99 & 0.2 & 1 \end{pmatrix}$$

$$Q = \left[\begin{pmatrix} 0.908 & 0.092 & 0 & 0 \\ 0.130 & 0.870 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.858 & 0.142 \\ 0 & 0 & 0.092 & 0.908 \end{pmatrix} \right]$$

FIG. 3.11. *A 3-D phase portrait showing accidental stability. The boundary of the simplex is indicated by glassy rods, and fixed points are indicated by metallic spheres. The figure's shadow is visible on the left.*

case of very similar UGs, evolutionary stability cannot be assured just by constraints on the payoff matrix. The initial conditions and learning algorithms of the different UGs can determine the outcome.

These mathematical results suggest that one should be especially careful when formulating hypotheses about the origins and genetic history of human language. Specifically, mutations leading to linguistic innovations must be reasonably compatible with the established language and UG to have any chance of taking over. Even then, historical accident and properties of the learning process influence the outcome. As the propositions and examples explored in this paper show, the relative fitness of one UG with respect to another is difficult to define, and any potential definition must be parameterized by the linguistic environment: Since UGs are metastrategies, they are one level removed from the payoff, so the particular mixture of languages present in a population may determine whether or not a mutation dies out or spreads.

This research could be extended in a number of directions. A discrete stochastic

model of a finite population might shed light on the question of how likely it is for a mutation in UG to spread. The results of Section 3 are independent of the learning algorithm, and give no indication of how the acquisition process might change over time. Thus, it would be informative to study cases of the model for two UGs that differ only in their learning algorithm. Also, the linguistic environment could be modeled in more detail, including features such as noisy linguistic data and social and spatial structure.

REFERENCES

[1] E. J. Briscoe, *Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device*, Language, 76 (2000), pp. 245–296.

[2] A. Cangelosi and D. Parisi, eds., *Simulating the Evolution of Language*, Springer-Verlag, 2001.

[3] Noam Chomsky, *Language and Problems of Knowledge*, MIT Press, 1988.

[4] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, 1990.

[5] M. D. Hauser, *The Evolution of Communication*, Harvard University Press, Cambridge, MA, 1996.

[6] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, 1998.

[7] J. R. Hurford, M. Studdert-Kennedy, and C. Knight, eds., *Approaches to the Evolution of Language*, Cambridge University Press, 1998.

[8] Ray Jackendoff, *Possible stages in the evolution of the language capacity*, Trends in Cognitive Sciences, 3 (1999), pp. 272–279.

[9] ———, *Foundations of Language*, Oxford University Press, Oxford, 2002.

[10] Simon Kirby, *Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity*, IEEE Transactions on Evolutionary Computation, 5 (2001), pp. 102–110.

[11] Natalia L. Komarova, Partha Niyogi, and Martin A. Nowak, *The evolutionary dynamics of grammar acquisition*, Journal of Theoretical Biology, 209 (2001), pp. 43–59.

[12] Natalia L. Komarova and Martin A. Nowak, *The evolutionary dynamics of the lexical matrix*, Bulletin of Mathematical Biology, 63 (2001), pp. 451–485.

[13] ———, *Natural selection of the critical period for language acquisition*, Proceedings of the Royal Society of London, Series B, 268 (2001), pp. 1189–1196.

[14] C. S. L. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, and A. P. Monaco, *A forkhead-domain gene is mutated in a severe speech and language disorder*, Nature, 413 (2001), pp. 519–523.

[15] David Lightfoot, *The Development of Language: Acquisition, Changes and Evolution*, Blackwell Publishers, 1999.

[16] W. Garrett Mitchener, *Bifurcation analysis of the fully symmetric language dynamical equation*, Journal of Mathematical Biology, 46 (2003), pp. 265–285.

[17] ———, *A Mathematical Model of Human Languages: The interaction of game dynamics and learning processes*, PhD thesis, Princeton University, 2003.

[18] W. Garrett Mitchener and Martin A. Nowak, *Competitive exclusion and coexistence of universal grammars*, Bulletin of Mathematical Biology, 65 (2003), pp. 67–93.

[19] ———, *Chaos and language*, Proceedings of the Royal Society of London, Biological Sciences, 271 (2004), pp. 701–704. DOI 10.1098/rspb.2003.2643.

[20] Partha Niyogi and Robert C. Berwick, *A dynamical systems model for language change*, Complex Systems, 11 (1997), pp. 161–204.

[21] P. Niyogi and R. C. Berwick, *Evolutionary consequences of language learning*, Linguistics and Philosophy, 20 (1997), pp. 697–719.

[22] Martin A. Nowak, Natalia L. Komarova, and Partha Niyogi, *Evolution of universal grammar*, Science, 291 (2001), pp. 114–118.

[23] ———, *Computational and evolutionary aspects of language*, Nature, 417 (2002), pp. 611–617.

[24] Martin A. Nowak and D. C. Krakauer, *The evolution of language*, Proceedings of the National Academy of Sciences, USA, 96 (1999), pp. 8028–8033.

[25] Martin A. Nowak, D. C. Krakauer, and A. Dress, *An error limit for the evolution of language*, Proceedings of the Royal Society of London, Series B, 266 (1999), pp. 2131–2136.

[26] Martin A. Nowak, Joshua Plotkin, and V. A. A. Jansen, *Evolution of syntactic communication*, Nature, 404 (2000), pp. 495–498.
[27] Martin A. Nowak, Joshua Plotkin, and D. C. Krakauer, *The evolutionary language game*, Journal of Theoretical Biology, 200 (1999), pp. 147–162.
[28] Steven Pinker, *The Language Instinct*, W. Morrow and Company, New York, 1990.
[29] Steven Pinker and A. Bloom, *Natural language and natural selection*, Behavioral and Brain Sciences, 13 (1990), pp. 707–784.
[30] Joshua Plotkin and Martin A. Nowak, *Language evolution and information theory*, Journal of Theoretical Biology, 205 (2000), pp. 147–159.
[31] Peter E. Trapa and Martin A. Nowak, *Nash equilibria for an evolutionary language game*, Journal of Mathematical Biology, 41 (2000), pp. 172–1888.